Ensemble Smoothing and Markov Chain Monte Carlo for Data Assimilation in Highly Nonlinear Systems

Michael Turmon (JPL)
T. M. Chin (JPL)
J. Jewell (JPL)
M. Ghil (ENS, UCLA)

For AGU, December 2005

Abstract

Current methods for atmosphere and ocean data assimilation propagate Gaussian distributions for gridded state variables forward in time. Powerful as these methods are, they do not handle outliers well and cannot simultaneously entertain multiple hypotheses about system state. The alternative of propagating the system's full probability density function (pdf) is computationally impractical.

Ensemble methods have been introduced into data assimilation for nonlinear systems as a compromise between the compact Gaussian parameterization and a gridded representation of the full pdf.

By propagating an ensemble of representative states, algorithms like the Ensemble Kalman Filter (EnKF) and the Resampled Particle Filter (RPF) rely on existing modeling infrastructure and capture the weights to be assigned to the state estimate based on the compatibility of an ensemble member with observed data.

We present an ensemble-based smoother that is applicable to Monte Carlo filtering schemes like the EnKF and the RPF. The algorithm does not require retrospective adaptation of the ensemble members themselves, and is thus suited to a streaming operational mode.

At the minor cost of retrospectively updating a set of weights for ensemble members, this smoother provides superior state tracking for two simple nonlinear problems, the double-well potential and the trivariate Lorenz system.

The accuracy of the proposed backward-update scheme in estimating non-Gaussian distributions is evaluated by comparison of its posteriors with ground truth provided by a Markov chain Monte Carlo (MCMC) algorithm.

Conclusions

- Inability to capture non-Gaussian posterior distributions is a handicap to conventional data assimilation practice.
- Particle-based methods – filters and smoothers – can overcome this limitation. Comparison with MCMC ground truth shows that particle methods can closely approximate the true posterior mean and variance in highly nonlinear systems.
- These ensemble-based methods are compatible with conventional forecasting data flows.
- The same smoothing methodology can be applied at minimal cost to reweight both EnKF and particle filter ensembles.
- Reweighting significantly reduces mean squared error in comparison with filters.

The research described in this poster was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (NASA).

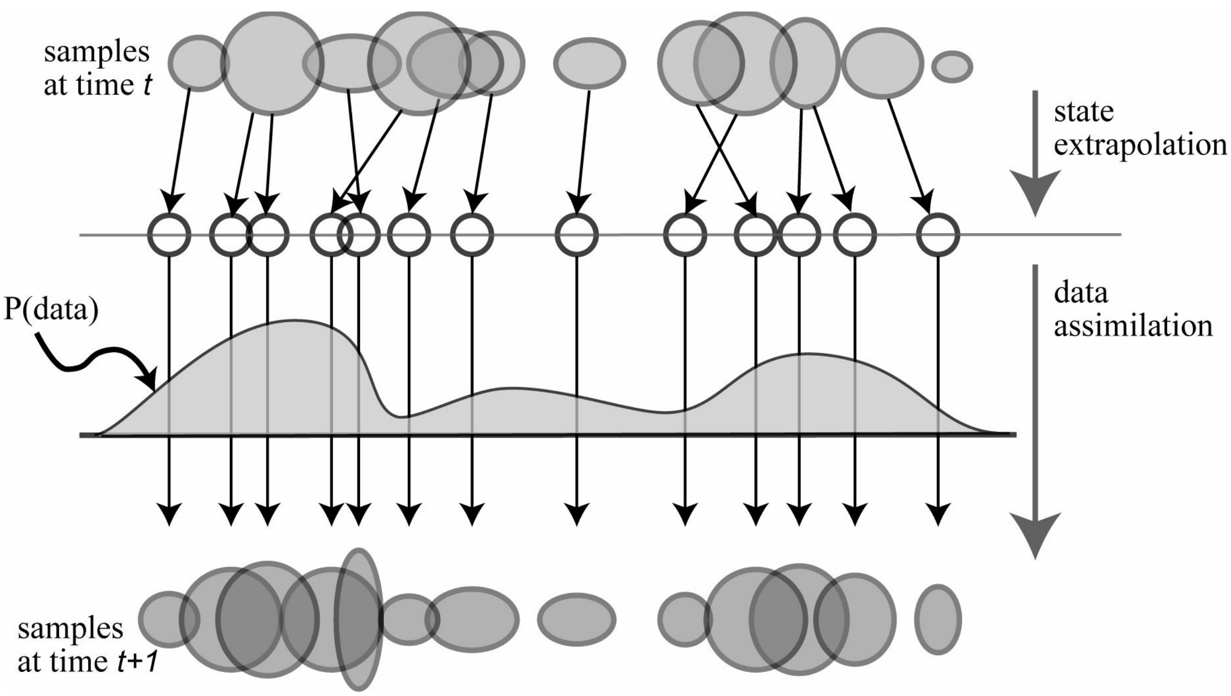
Ensembles for Data Assimilation

Approaches for data update (Bayes' Rule):

Ensemble Kalman filter (EnKF) where Gaussian distribution is assumed.

Gaussian mixture where the distribution is sum of multiple Gaussian functions to handle multi-modal dynamics.

Particle filter where each sample is its value and probability $\left\{\mathbf{x}_t^{(n)}, p_t^{(n)}\right\}$



Horizontal: state value. Vertical: relative probability.

Particle Smoother

We seek the conditional pdf $p(\mathbf{x}_t|\mathbf{Y}_u)$ of the atmosphere/ocean state $\mathbf{x}_t$ given data $\mathbf{Y}_u = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_u\}$,

For filtering, use $u = t$; for smoothing, set $u = T$.

Particle filter allows time-sequential computation of $p(\mathbf{x}_t \,|\, \mathbf{Y}_t)$, which is approximated by the N weighted samples $\left(\mathbf{x}_t^{(n)}, p_{t|t}^{(n)}\right)$, $1 \leq n \leq N$

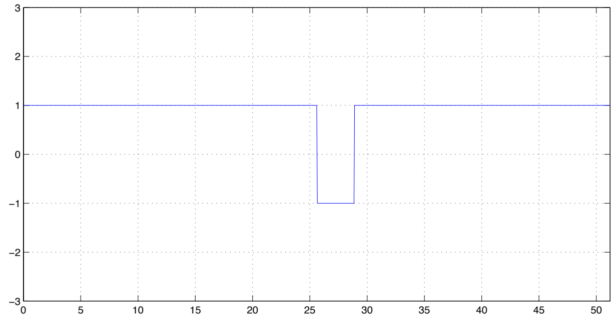
Particle smoother rescales the weights $p_{t|t}^{(n)}$ to $p_{t|T}^{(n)}$, leaving the corresponding state samples $\mathbf{x}_t^{(n)}$ untouched. The smoother weight update accounts for the forward-looking dynamics.

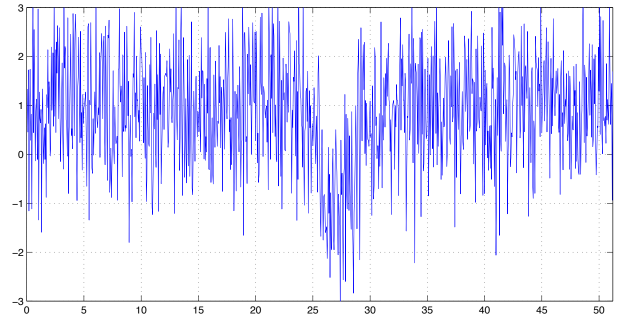
Key computational step reweights $p_{t|T}^{(n)} = c_t \, p_{t|t}^{(n)}$ where c_t is a function of $p_{t+1|T}$.

$p_{t|T}^{(n)}$ can be computed backwards-sequentially as a series of matrix multiplications on the weights.

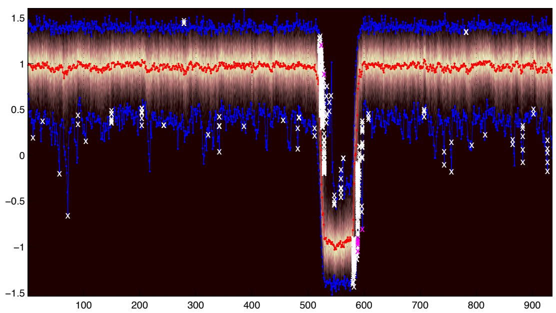
Particle Filter: Usage Example

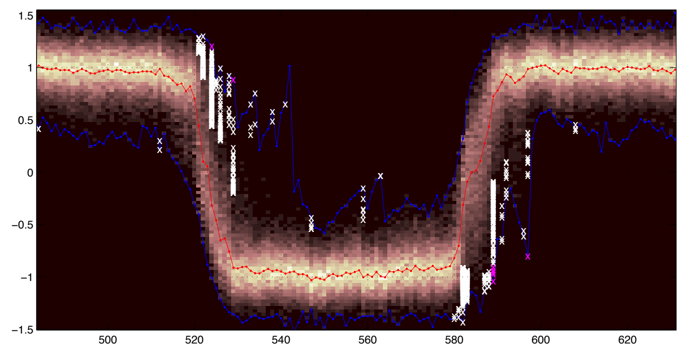
True $x_1 ... x_T$ with data (noise $\sigma = 1$)





Main limitation is loss of diversity of particle set under fast transitions





MCMC: Essential Idea

MCMC: A general way to approximate
expectations of any probability distribution

Specifically: Posterior mean, posterior clusters, or
the posterior probability density itself

For now, abbreviate $p(x_t \,|\, y_1, \ldots, y_T)$ by $\pi(x)$

- **MCMC is Monte Carlo integration...**

$$Ef(X) = \int f(x)\,\pi(x)\,dx \simeq \frac{1}{N}\sum_{i=1}^{N} f(x_i) \quad (x_i \text{ from } \pi)$$

Typically x_i are sampled independently but the
approximation works even for dependent x_i

- **...driven by Markov chain samples**

Craft a MC having $\pi(\cdot)$ as its stationary dist'n
Draw the x_i's above from this MC
They are not independent, but they are from π

The sample average of $f(x_i)$ converges to $Ef(X)$,
same as ensemble averages do (ergodicity)

MCMC: Metropolis-Hastings

One particular kind of MCMC uses the...

- **Metropolis-Hastings recipe**

  At some simulation epoch, the MC is at x
  Draw a candidate x' from $q(x' \,|\, x)$
  Replace x by the new x' with probability

  $$\rho = \min\Big(1, \frac{\pi(x')\, q(x \,|\, x')}{\pi(x)\, q(x' \,|\, x)}\Big)$$

  Under broad conditions, sample mean $\to Ef(X)$
  (rate of convergence is hard to assess)

  Can use un-normalized versions of $\pi(\cdot)$

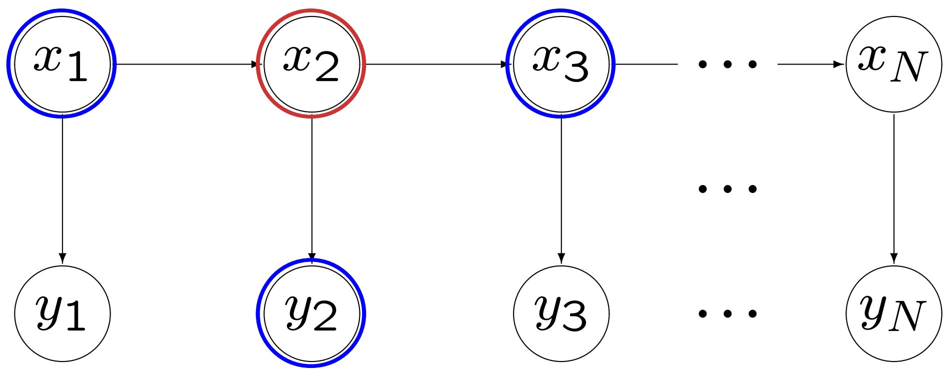
- **Acceptance probability ρ**

  Favors increasing the posterior, but allows $\pi(x)$ to decrease occasionally

  Adjusts for any bias introduced by our choice of q

MCMC: Applied to Time Series

The Bayes network for the time series shows dependences among variables:



For now, suppose we only want to estimate x_2 and other values are known

Apply Metropolis-Hastings recipe:

Propose a new value x_2', e.g. from $N(x_2, \delta\kappa^2)$
($q(x_2' \,|\, x_2)$ is symmetric)

Compute the ratio

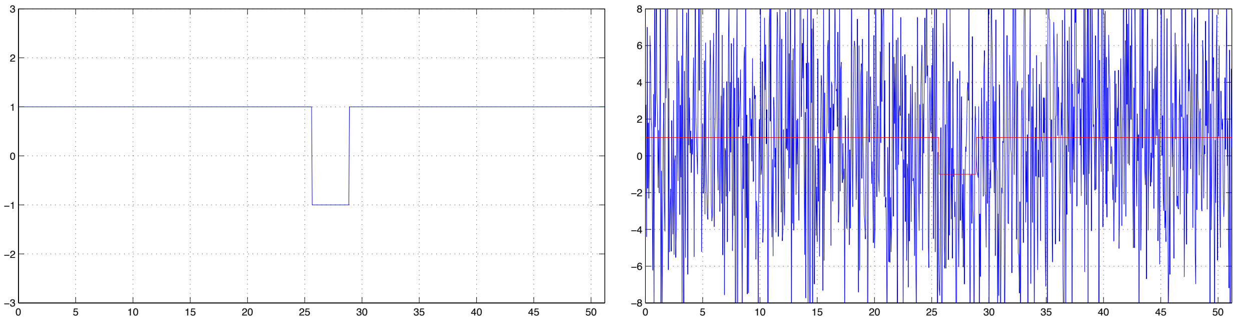
$$\frac{\pi(x')\, q(x \,|\, x')}{\pi(x)\, q(x' \,|\, x)} = \frac{p(x_2' \,|\, x_1)\, p(x_3 \,|\, x_2')}{p(x_2 \,|\, x_1)\, p(x_3 \,|\, x_2)} \times \frac{p(y_2 \,|\, x_2')}{p(y_2 \,|\, x_2)}$$

Combines a smoothness term and a data term

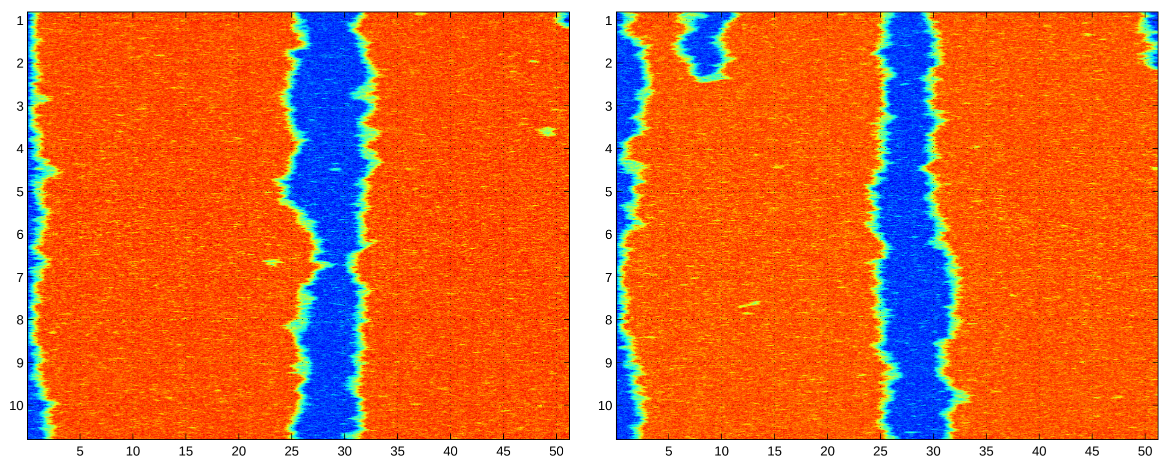
The full MCMC scheme sweeps over all unknown variables $x_1, \ldots, x_T$, proposing changes to each

MCMC: Usage Example

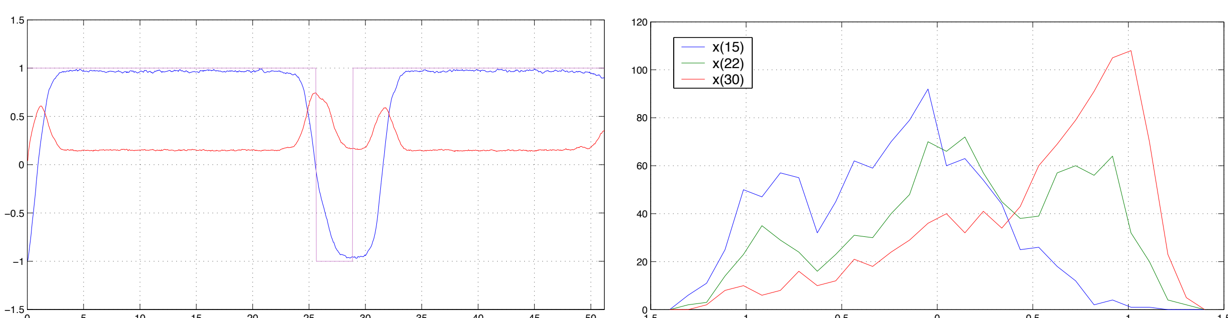
True $x_1 \ldots x_T$ with data (obs. noise $\sigma = 5$)



Two MCMC series (10^4:10^5 downsampled 100:1)



$E\, x_t$, $\sigma(x_t)$, truth; $p(x_t \,|\, y_{1:T})$, $t = 0.75, 1.1, 1.5$

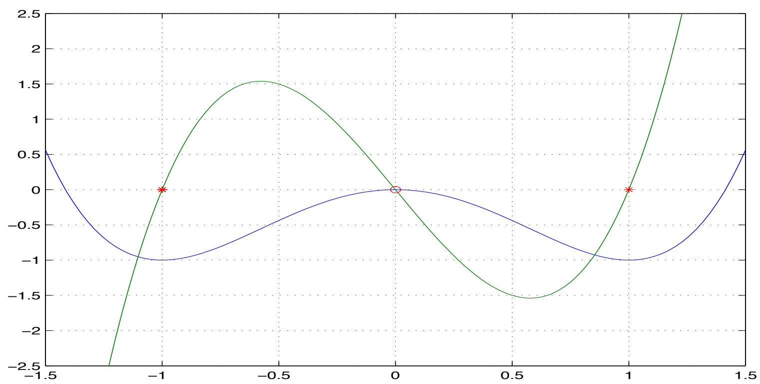


Note: skew distribution at $t = 1.5$, multimodal distribution at $t = 1.1$

Dynamics: Double-Well aka DW

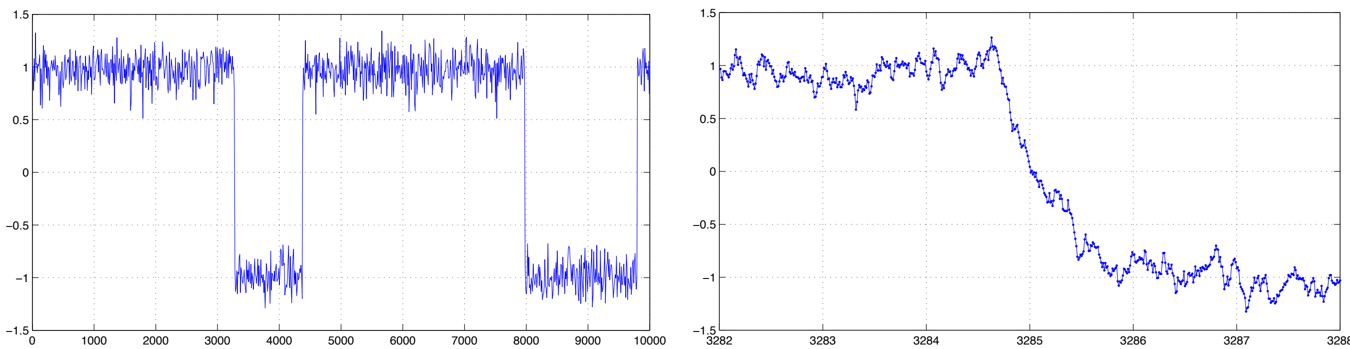
Particle in a potential well given by
$f(x) = -2x^2 + x^4$ and $g(x) = (d/dx)f(x)$

Two minima at ± 1 and one stationary point at 0



Trivial climate: Miller et al., Eyink & Restrepo, ...

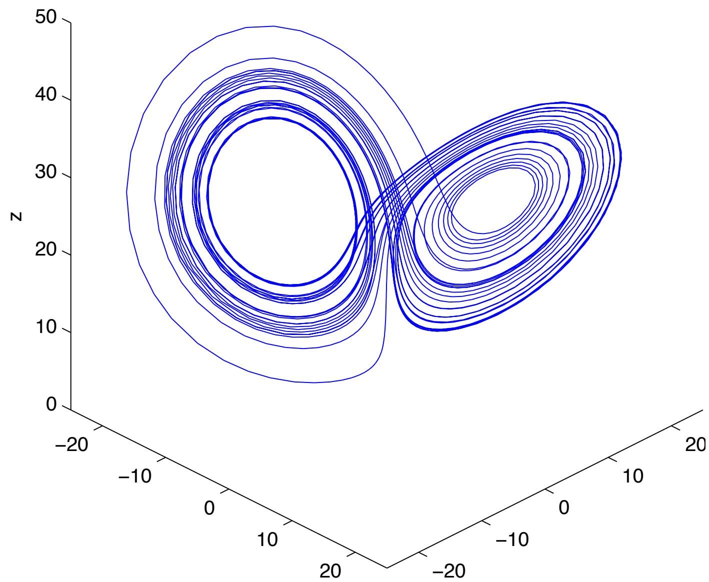
Observations cluster at ± 1, occasionally switching state: highly non-Gaussian



Dynamics: Lorenz 1963 aka L63

We also use the trivariate Lorenz 1963 system:

$$\dot{x_1} = 10(x_2 - x_1)$$
$$\dot{x_2} = 28\,x_1 - x_2 - x_1 x_3$$
$$\dot{x_3} = x_1 x_2 - (8/3) x_3$$



In vector form, $\dot{\mathbf{x}}_t = -g(\mathbf{x}_t)$ for g as above

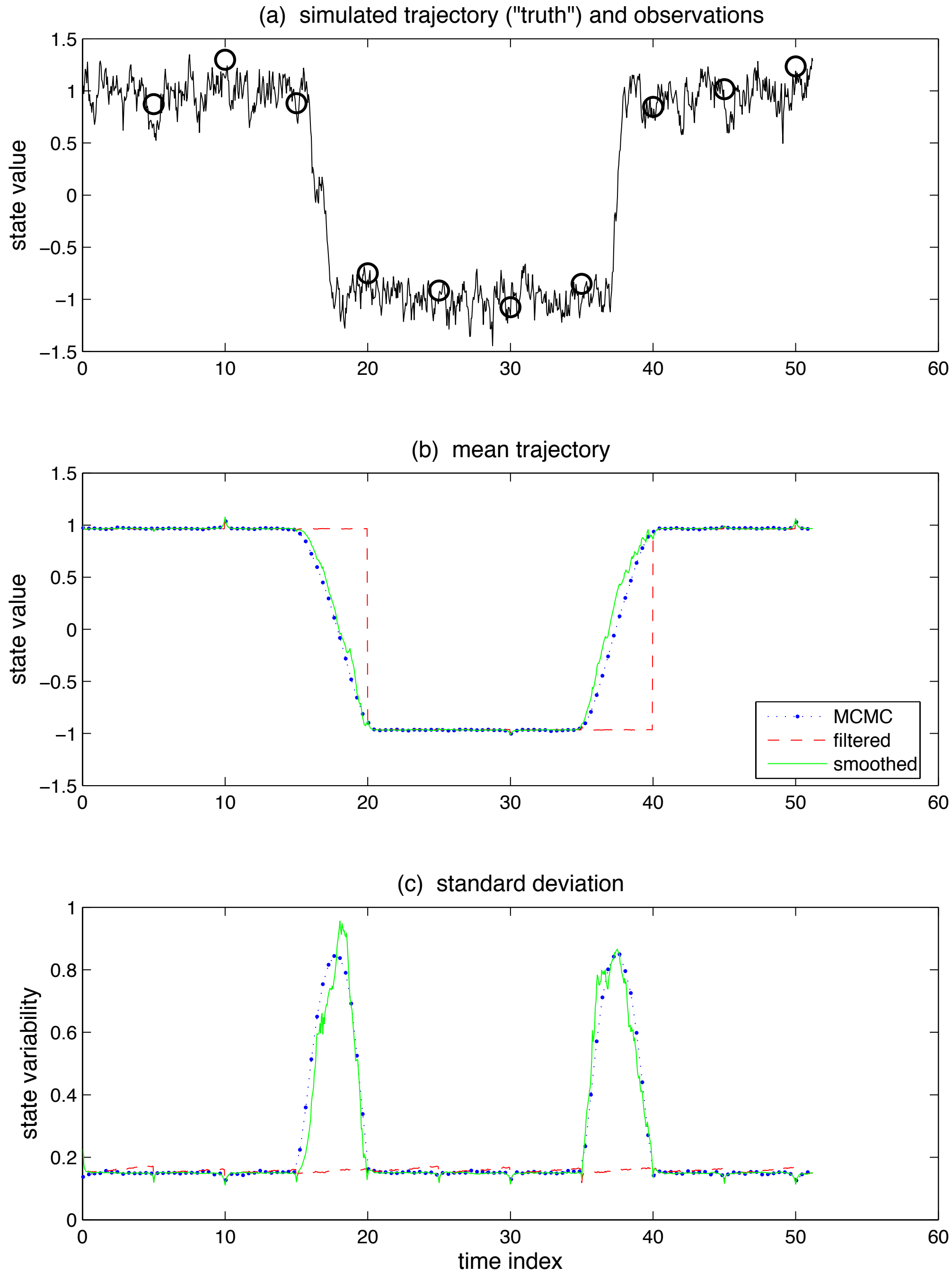
For experiments, discretize DW and L63 in time:

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \Delta \mathbf{g}(\mathbf{x}_{t-1}) + \mathbf{w}_t \qquad \mathbf{w}_t \sim \mathcal{N}(0, \kappa^2 \Delta \mathbf{I})$$
$$\mathbf{y}_t = \mathbf{x}_t + \mathbf{v}_t \qquad \mathbf{v}_t \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$$

Notation "$\sim \mathcal{N}(\mu, \mathbf{R})$" means "is Gaussian distributed with mean vector μ and covariance matrix $\mathbf{R}$"

$\mathbf{y}_t$ is the noisy observation

DW: Trajectory Estimates

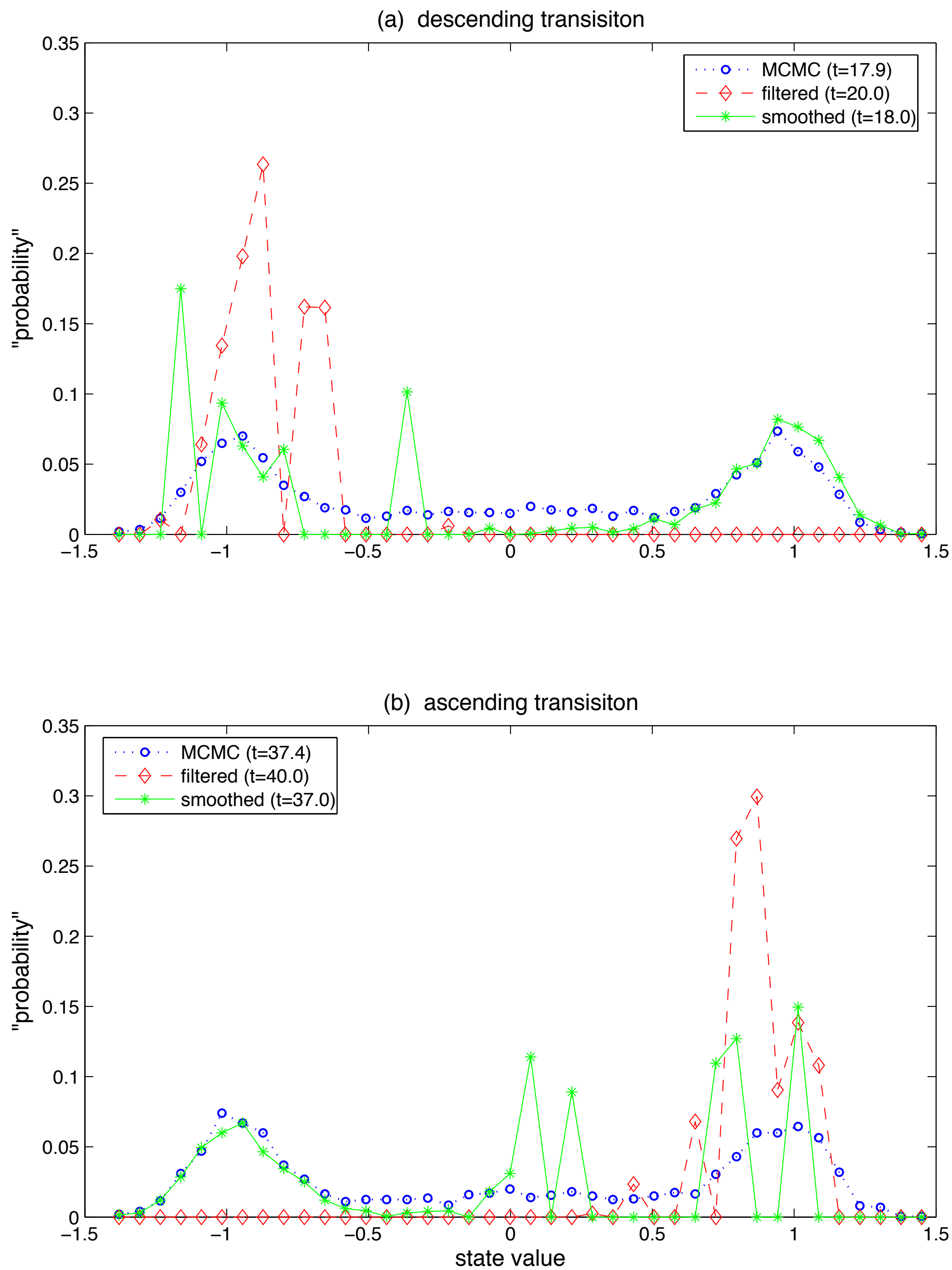


Top: DW trajectory with sparse, noisy observations (circles).

Middle: Mean trajectories from MCMC (2000 samples), particle filter ($N = 10^4$ particles), particle smoother

Bottom: Standard deviations $std(\mathbf{x}_t)$ given the observations

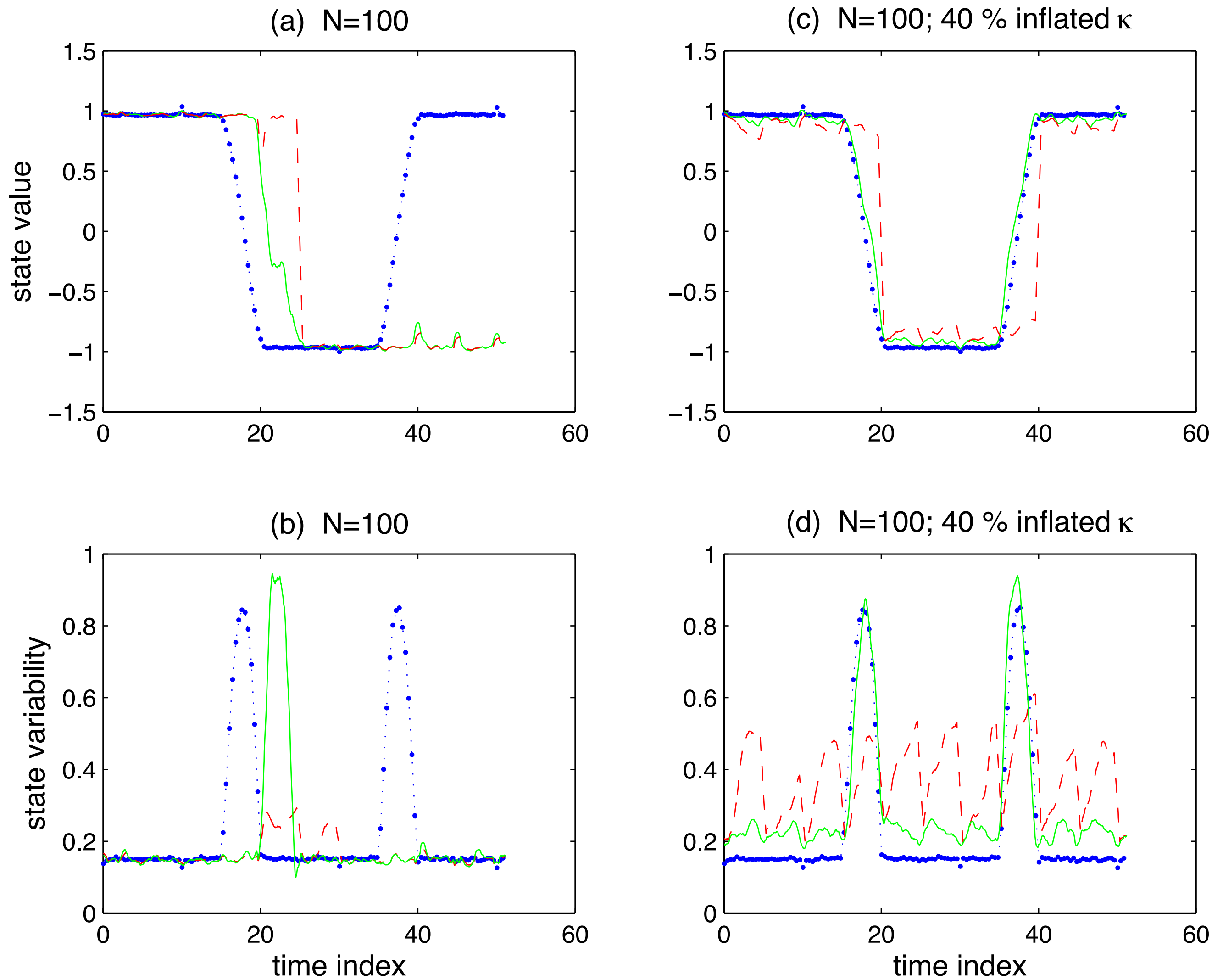
DW: Nongaussian Posterior



Normalized histograms of the state values $p(\mathbf{x}_t \,|\, \mathbf{Y}_t)$ at the two zero-crossings

Note strongly nongaussian posteriors captured by MCMC and particle methods

DW: Small Ensembles

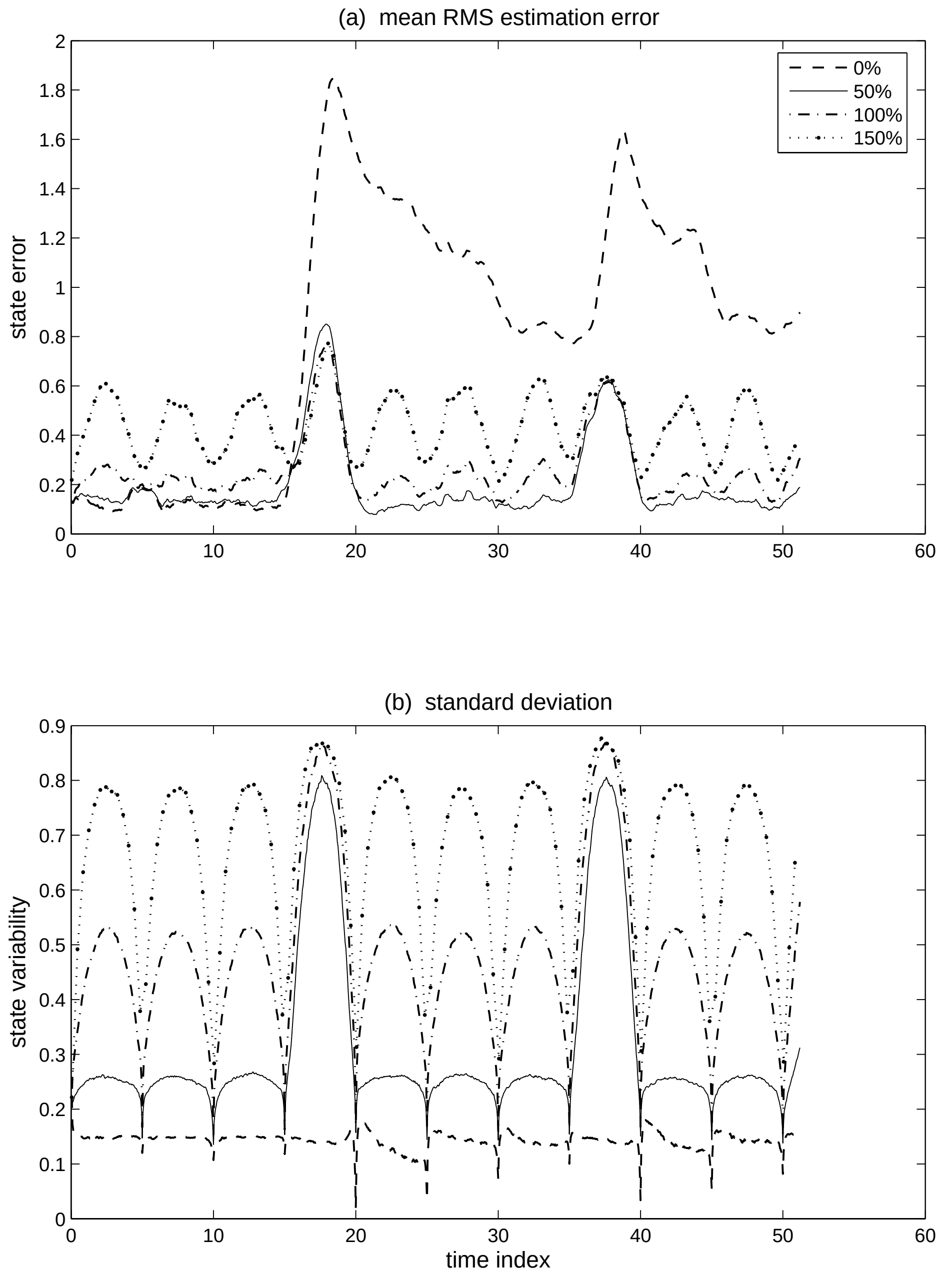


Left: Same as above, using only $N = 100$ particles. Filter and smoother tracking degrades relative to MCMC ground truth.

Right: $N = 100$, but the system noise κ is artificially inflated by 40%, restoring performance.

This is one way to shrink the number of samples for better computational characteristics

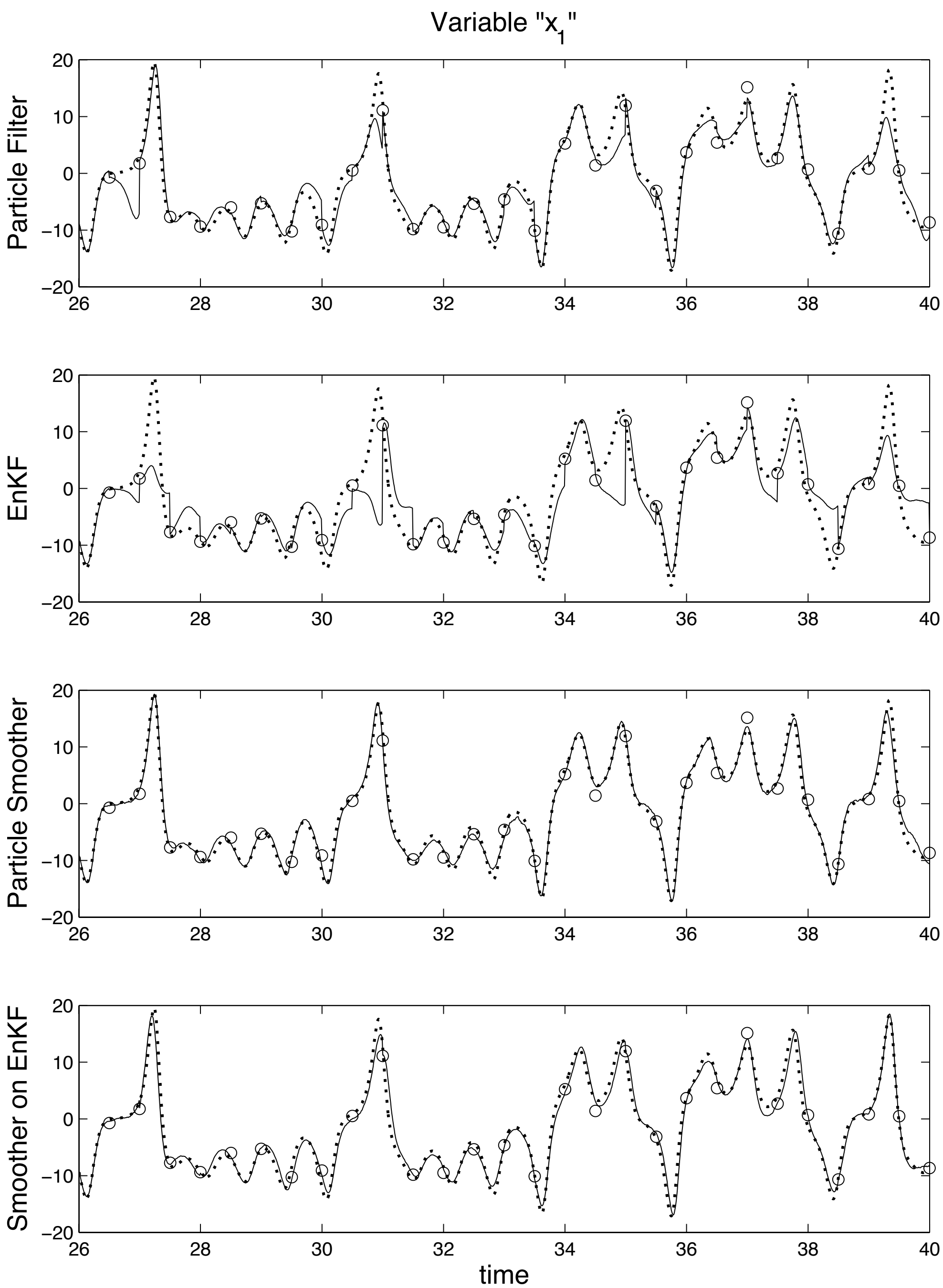
DW: Fixes for Small Ensembles



Top: Estimation error for the smoother, varying assumed system noise level κ. Best κ is 50%.

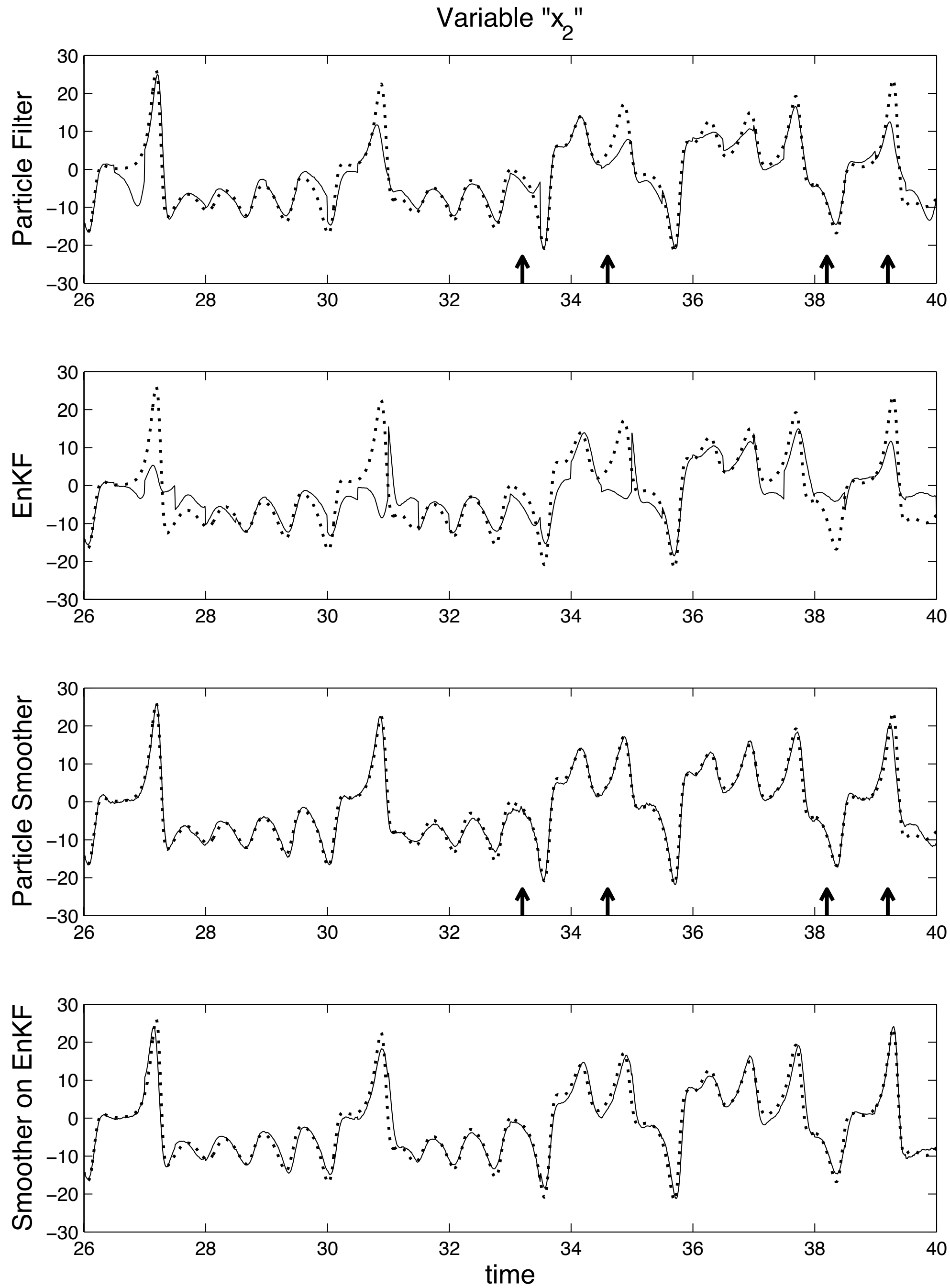
Bottom: Conditional variance $std(\mathbf{x}_t)$ at this κ preserves many correct characteristics relative to the $N = 10^4$ ensemble above.

L63: Particle Smoother Results



Each panel shows the L63 x_1 variable (dotted line) and its noisy observations (circles).

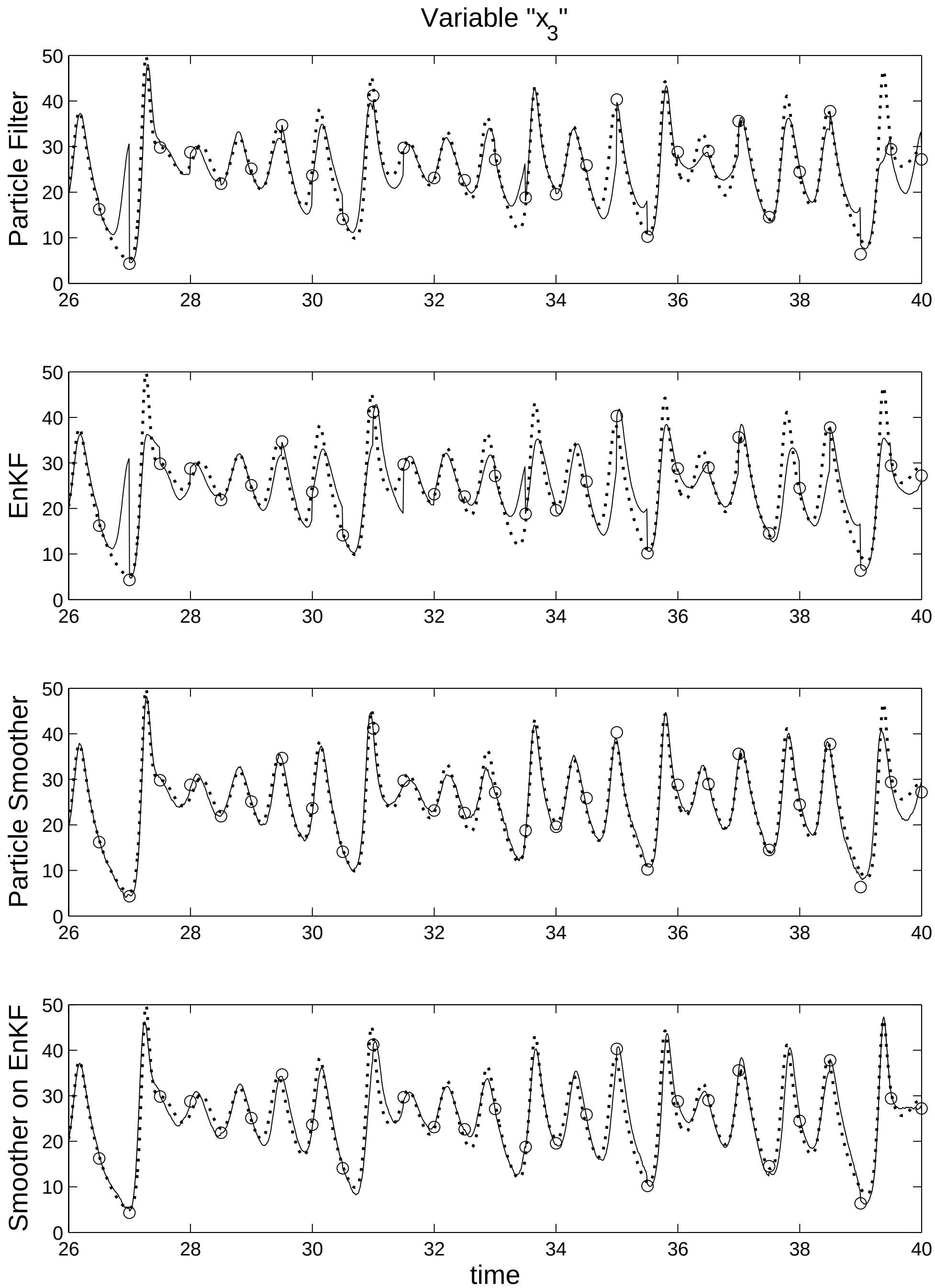
Solid lines show (from top to bottom) estimates by RPF, EnKF, RPF-BSS, and EnKF-BSS.



As above, but showing the x_2 variable, which is not directly observed.

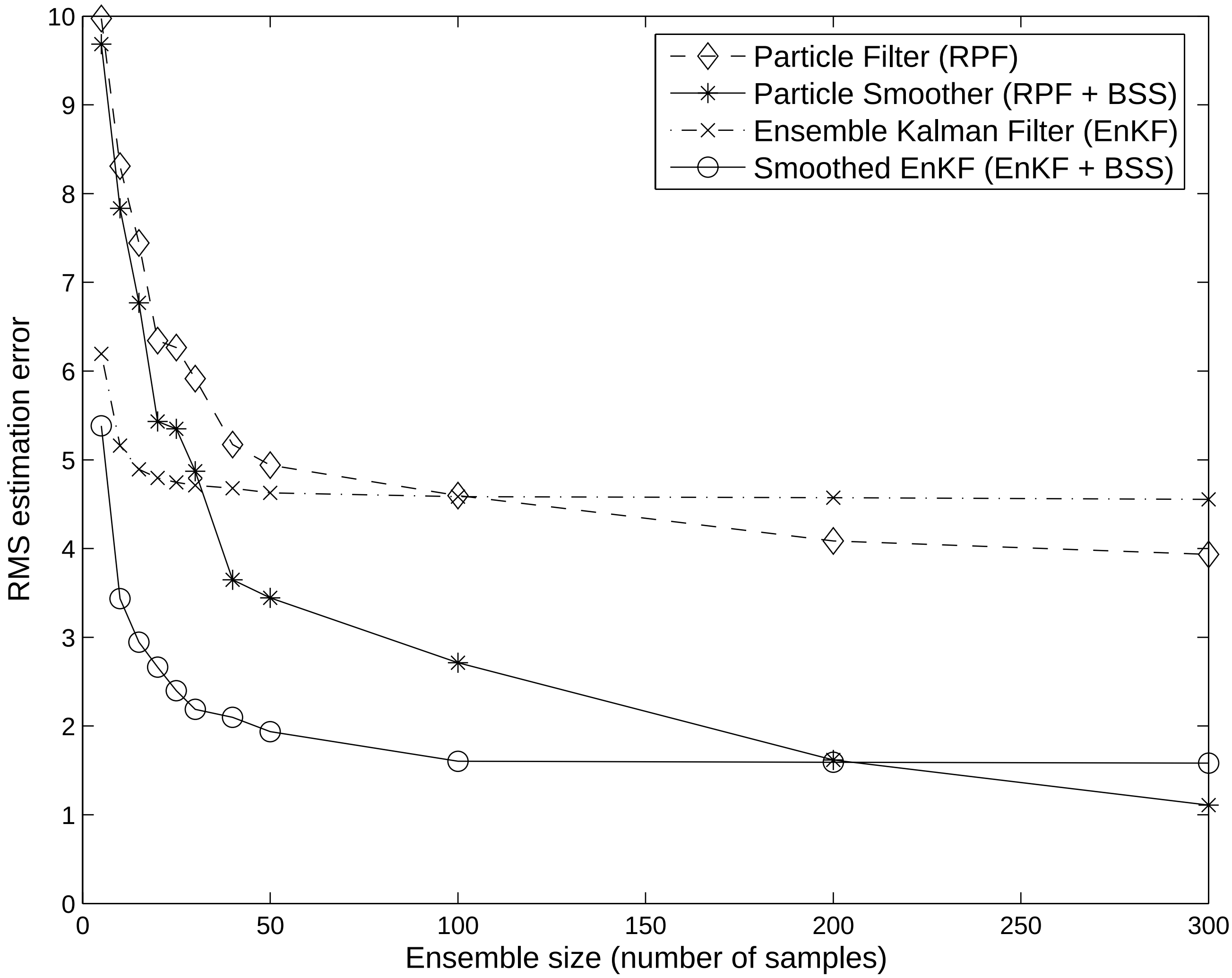
The particle filter (resp. smoother) significantly outperforms the EnKF filter (resp. smoother) in this highly nonlinear problem.

Furthermore, the ensemble-based smoothers have considerably better tracking than the filters.



As above, showing x_3.

L63: Ensemble Size and Bias



RMS estimation errors (averaged over 50 runs) for L63 as ensemble size is varied.

For ensemble size below 20, EnKF outperforms the nonparametric approaches (PF/PS).

Increasing ensemble size beyond ∼30 does not improve EnKF accuracy.

This shows the price paid in estimator bias for making the Gaussian assumption.